STOCK MARKET PREDICTION USING LSTM

TEAM MEMBERS:
1) IRFAN C [au723721243021]
2) KUMARAGURU SV[au723721243029]
3) LINGESHWARAN J[au723721243030]
4) ARIVUMATHY M[au723721243007]
5) ROSHAN KUMAR S [au723721243042]

PHASE 5 PROJECT REPORT

Problem Statement:
Predicting stock price using Microsoft life time stock price dataset by applying learning algorithms like ARIMA and LSTM.

DESIGN THINKING PROCESS:

Empathy:
Investors rely on stock price predictions to make informed decisions about buying, holding, or selling stocks. Investors want accurate predictions to minimize risk and maximize returns. Predictions should align closely with actual stock price movements. They expect insights into whether a stock is in an uptrend, downtrend, or consolidation phase. Real-time news and event analysis can be vital, especially for short-term traders. Transparent and interpretable models can help investors understand why a particular prediction was made, fostering trust in the prediction system.

Definition: -
The project's core objective is to create a predictive model for stock price forecasting using historical market data, primarily to aid investors in decision-making and strategy optimization. This endeavor involves crucial stages, including data collection, pre processing, feature engineering, **Stock market prediction** model selection, training, and evaluation.

Predicting stock prices is intricate due to factors like non-stationary data, noise, poor model generalization, unexpected "Black Swan" events, and the impact of regulatory changes on market dynamics, necessitating model adjustments.

Idea:
Utilize historical stock price data to create time series models like ARIMA or Seasonal Decomposition of Time Series (STL) to make short-term predictions and models like Prophet or LSTM networks for longer-term predictions. Apply

machine learning algorithms like linear regression, decision trees, random forests, or gradient boosting to predict stock prices. Analyse a company's financial statements, earnings reports, and economic indicators to assess its intrinsic value and make long-term predictions. Monitor social media platforms, financial news, and online forums to gauge market sentiment and investor opinions. Develop a simulation framework to test the performance of various prediction models and trading strategies over historical data.

Design:
Analyzing stock market data using time series analysis techniques like ARIMA (Auto Regressive Integrated Moving Average) and LSTM (Long Short-Term Memory) can provide valuable insights and help make predictions.

Data Source:
A good data source for prediction using deep learning should be Accurate, Complete, Covering the geographic area of interest, Accessible.

Dataset Link: (https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset)

The dataset contains several variables, including date, open, high, low, close, and volume. The columns Open and Close represent the opening and closing prices of the stock on a given day. The maximum and minimum share prices for the day are represented by High and Low. The number of shares purchased or sold during the day is referred to as volume. Another thing to keep in mind is that the market is closed on weekends and public holidays.

Data preprocessing:

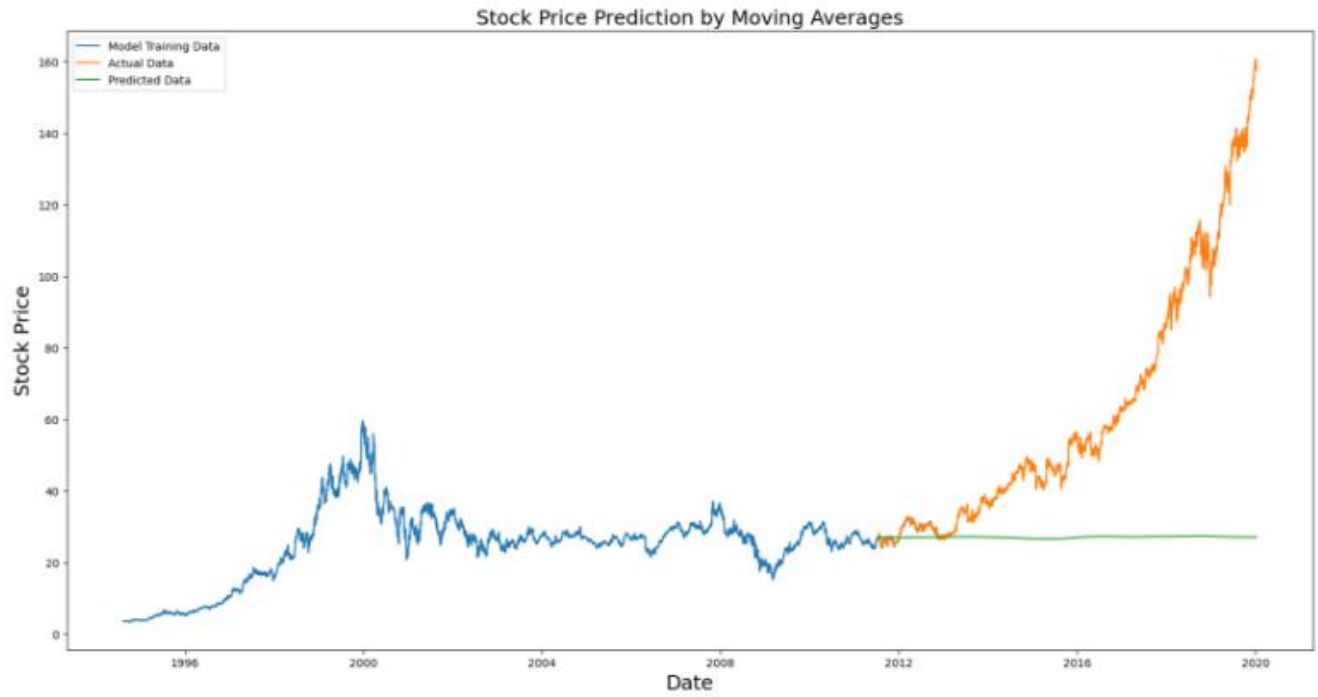
- The data is checked for null values, number of null values = 0
- The data is checked for irrelevant data, number of irrelevant data = 0
- The duplicates in the dataset are dropped.
- The ‘date’ column is converted into date time column in excel and the action is completed with pandas (from YYYY-MM-DD to MM/DD/YYYY HH:MM:SS)

	Date	Open	High	Low	Close	Adj Close	Volume
0	03/13/1986 00:00:00	0.088542	0.101563	0.088542	0.097222	0.062549	1031788800
1	03/14/1986 00:00:00	0.097222	0.102431	0.097222	0.100694	0.064783	308160000
2	03/17/1986 00:00:00	0.100694	0.103299	0.100694	0.102431	0.065899	133171200
3	03/18/1986 00:00:00	0.102431	0.103299	0.098958	0.099826	0.064224	67766400
4	03/19/1986 00:00:00	0.099826	0.100694	0.097222	0.098090	0.063107	47894400

MODEL TRAINING PROCESS:

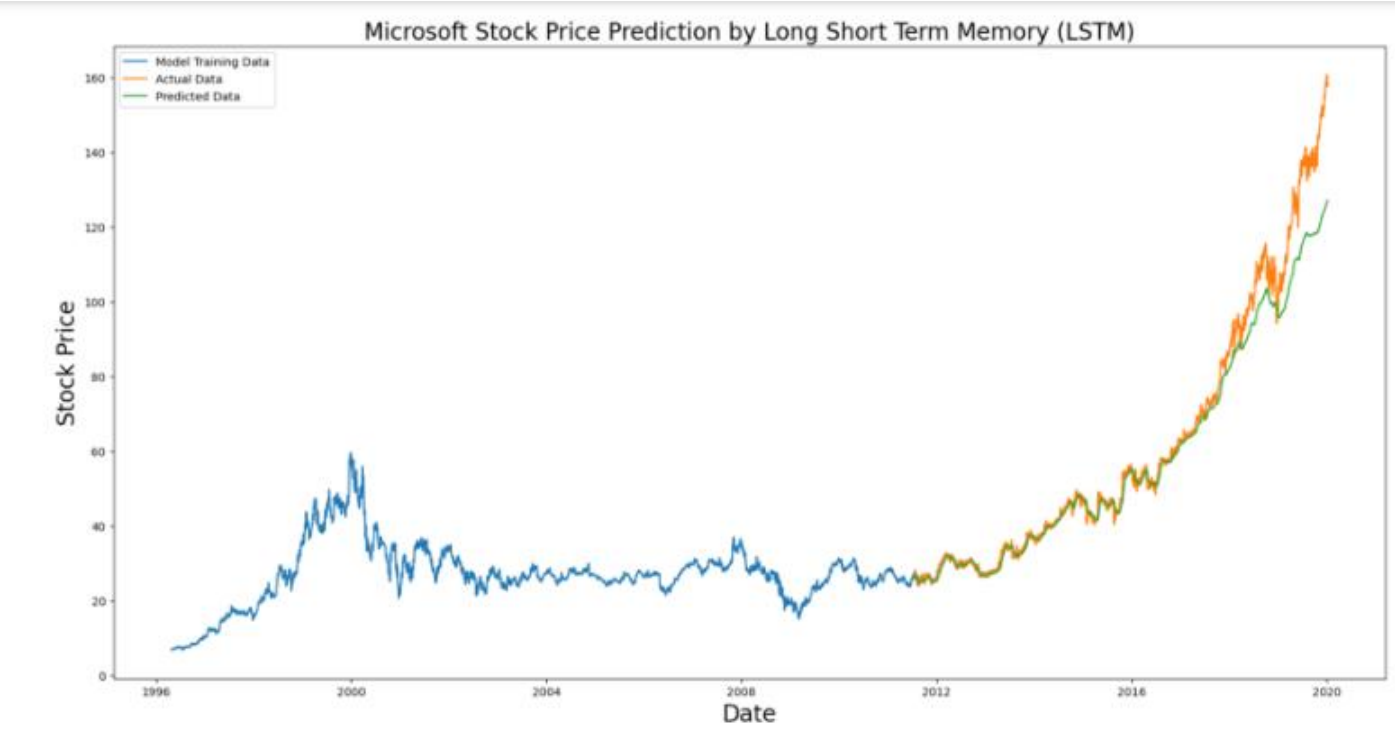
ARIMA MODEL:

- The dataset is split into X axis and Y axis, Y axis consist the 'Close' column, the 'date' column is taken as index.
- The dataset is split into train and test dataset and it has been reshaped.
- The Moving average is done by mathematically calculating the moving average of the training data.
- It is validated by using the test data.
- The performance metrices has been found and the prediction is plotted along with the actual values.



LSTM MODEL:

- The dataset is split into X axis and Y axis, Y axis consist the ‘Close’ column, the ‘date’ column is taken as index.
- The dataset is split into train and test dataset and it has been reshaped.
- The sequential model is initiated and LSTM model with 50 units is added.
- To the model the dense neural network layer is added.
- The model is fitted into the training data with single epoch and batch size of 2.
- The test data has been reshaped and used for validation of the model.
- The performance metrices has been found and the prediction is plotted along with the actual values.



OBSERVATION:

Moving average performance metrices :

RMSE:48.88457740518917

r2:-1.0234099920313238

LSTM performance metrices :

RMSE:7.440627

r2:-82.5785420245649

INSIGHTS:

From the observation we can see that the Root Mean Square Error of LSTM(7.440627) is low compared to Root Mean Square Error of Moving average (48.88457740518917). Thus we can say that LSTM predicts much accurately compared to Moving average. The r2 value of moving average represents the linearity of the prediction. The r2 value of LSTM represents the fluctuations in a prediction.

Conclusion :

The Time Series Forecasting of Microsoft stock price prediction is implemented two sequential models namely Moving average and LSTM. Due to the averaging function of moving average model it fails to capture the increasing trend of recent time stock price data. LSTM model captures the correlation of the data much accurately.

Future Scope:

The stock market depends on various parameters other than the daily stock prices like statements made by the executives, promotions, product releases, etc., by including these factors a well accurate model can be developed.